

## DISTRIBUIÇÃO DE PROBABILIDADE:

### curva normal e aplicação

Aline Rosa Rodrigues

Discente do Curso de Licenciatura em Matemática – Uni-FACEF

alinerosarodrigues12@gmail.com

Bianca Cecília Silva

Discente do Curso de Licenciatura em Matemática – Uni-FACEF

biancaaliberti19@gmail.com

Letícia Faleiros Chaves Rodrigues

Mestre em Matemática Universitária e Docente do Uni-FACEF

leticia@facef.br

## RESUMO

Cada vez mais, o número de informações e o volume de dados crescem de uma forma grandiosa, sendo necessário pessoas e técnicas estatísticas capazes de analisar e compreender estes dados, e assim tomar decisões eficientes com qualidade. Desta forma, o objetivo do presente trabalho é realizar um estudo teórico da estatística descritiva, probabilidade e ferramentas estatística para processamento e análise de dados, apresentado conceitos e definições, abordando medidas de centralidade e dispersão, além de explanar quatro distribuições de probabilidade, sendo elas a Binomial, de Poisson, Exponencial e a Normal. Por fim, devido a distribuição normal ser amplamente usada, pois através dela, é possível modelar uma infinidade de fenômenos naturais, aproximar probabilidades de variáveis que possuem outras distribuições, optou-se por focar na mesma, a fim de compreendê-la e aplicá-la. No período de 2019 à meados de 2022 todo o mundo passou por uma pandemia denominada COVID-19, em que, desde seu início, muitos estudos e pesquisas foram realizados, assim foi apresentada uma aplicação da distribuição normal na análise do comportamento da série temporal do número de casos de pessoas contaminadas por dia. Deseja-se que a pesquisa e a aplicação possam contribuir de forma significativa para a compreensão do estudo.

**Palavras-chave:** Estatística. Probabilidade. Distribuições. Gauss. COVID-19.

## ABSTRACT

More and more, the number of information and the volume of data grow in a great way, requiring people and statistical techniques capable of analyzing and understanding this data, and thus making efficient decisions with quality. Thus, the objective of this paper is to conduct a theoretical study of descriptive statistics, probability and statistical tools for processing and analyzing data, presenting concepts and definitions, addressing measures of centrality and dispersion, in addition to explaining four probability distributions, namely Binomial, Poisson, Exponential and Normal. Finally, due to the fact that the normal distribution is widely used, because through it, it is possible to model a multitude of natural phenomena, approximate probabilities of variables that have other distributions, it was decided to focus on it, in order to understand and apply it. Currently, we are experiencing a pandemic called COVID-19, where, since its inception many studies and research have been carried out, so it was presented an application of the normal distribution in the analysis of the behavior of the time series of the number of cases of contaminated people per day. We consider that the research and application contributed significantly to the understanding of the study.

**Palavras-chave:** Statistics. Probability. Distributions. Gauss. COVID-19..

## 1 INTRODUÇÃO

A estatística desempenha um papel importante na sociedade moderna, pois fornece ferramentas para analisar variáveis de diferentes perspectivas, verificar possíveis relações por meio de pesquisas e experimentos e, assim, fornecer um caminho coerente e direto para possíveis decisões.

A estatística descritiva é um campo de estudo que se concentra na produção de um método de coleta, organização, descrição, análise e interpretação de dados, como tirar conclusões válidas e tomar decisões sensatas com base nesta análise. Os professores analisam dados dos alunos; o analista faz algumas análises sistemáticas de dados dos sistemas de informação; os médicos analisam os exames e as respostas dos pacientes ao tratamento e todos nós vemos estatísticas de censos, pesquisas eleitorais, bolsas de valores e assim por diante. Ela pode ser aplicada em praticamente todas as áreas do conhecimento humano, tais como, Economia, Indústria, Comércio, Educação, Administração, Agricultura, Informática, Farmácia, Psicologia, Medicina e várias outras.

Um experimento, que ao ser realizado sob as mesmas condições e não produz os mesmos resultados, é denominado um experimento aleatório, pois não podemos dar um parecer antes do término do experimento.

Esses fenômenos ocorrem frequentemente no nosso cotidiano, por exemplo: quem vai ganhar a Mega Sena? Choverá daqui uma semana? Lançamento de uma moeda honesta para cima, qual a face que ficará voltada para cima? Ainda que o resultado de um experimento aleatório não possa ser previsto, é possível conhecer todos os resultados possíveis para o experimento. O conjunto de todos os possíveis resultados desse experimento é denominado espaço amostral, que é indicado pela letra grega ômega ( $\Omega$ ). Pode conter um número finito ou infinito de pontos. Exemplo: {cara, coroa}, entre outros.

Na teoria da probabilidade, as distribuições descrevem o comportamento aleatório de fenômenos que dependem do acaso. O estudo dos fenômenos aleatórios começa com o estudo dos jogos de azar - jogos de dados, sorteio de bolas de urna, cara ou coroa, são motivações para entender e prever experimentos aleatórios. Esses métodos iniciais são fenômenos discretos, o que significa que o número de resultados possíveis é finito ou contável. No entanto, certos problemas revelam distribuições de probabilidade com incontáveis resultados.

Por exemplo, quando o lançamento de uma moeda vai ao infinito, o número de coroas se aproxima de uma distribuição normal, pois quando um experimento ultrapassar trinta resultados, já podemos usá-la.

Oscilações e variações existem em quase todos os valores que podem ser medidos durante a observação de um fenômeno, independentemente de sua natureza e, além disso, quase todas as medições possuem uma parte inerente de erro. A distribuição de probabilidade pode modelar incertezas e descrever fenômenos físicos, biológicos, econômicos, entre outros. O domínio da estatística permite o encontro das distribuições de probabilidade adaptadas aos fenômenos aleatórios.

O objetivo deste trabalho é discorrer sobre quatro distribuições de probabilidade: Normal, Binomial, Poisson e Exponencial, aplicando em diferentes situações que ocorrem no cotidiano. Desta maneira o presente trabalho foi assim dividido.

No primeiro capítulo, apresentamos análise descritiva, onde será abordado sobre tipos de variáveis os quais podem ser discretas ou contínuas, pois a compreensão das variáveis se faz necessário para buscar identificar a melhor técnica a ser utilizada e assim desenvolver o modelo ideal para o estudo em questão. Será explicado de forma breve sobre as medidas de centralidade, o qual nada mais é do que um centro de distribuição de dados, as quais são: média, moda e mediana, bem como sobre medidas de dispersão, que é utilizada para que se possa compreender qual é o grau de variação dos dados examinados, porque afinal de contas, nenhuma medida é realmente exata, já que há inúmeros fatores que influenciam a coleta de dados.

O segundo capítulo trata da teoria da probabilidade, recordando brevemente o cálculo de probabilidade necessário para entender esta pesquisa, na qual aborda-se: conjunto, espaço amostral, eventos e probabilidade. O terceiro capítulo discorre sobre as distribuições de probabilidades, tanto as discretas, na qual será trabalhado sobre distribuição binomial e distribuição de Poisson, como também, as distribuições contínuas, em que será discorrido a distribuição exponencial e a distribuição normal. Por fim, no último capítulo, será realizada uma aplicação utilizando a distribuição normal para analisar os dados relacionados à pandemia COVID19.

RELMAT

Revista Eletrônica do Curso de Licenciatura em Matemática

Uni-FACEF

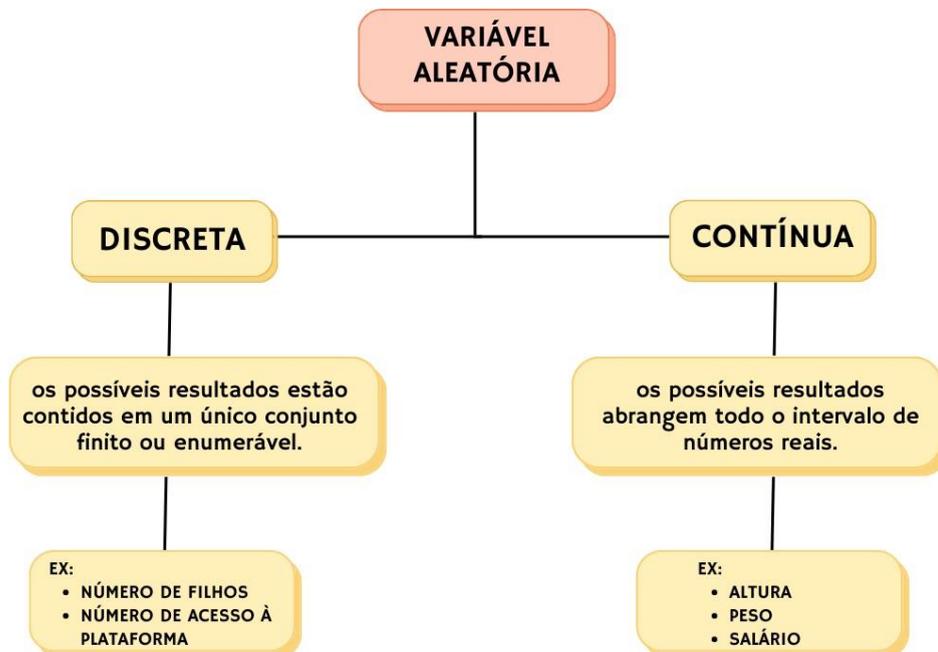
Centro Universitário Municipal de Franca

Aborda a sobre as pandemias ao longo da história e analisa os dados obtidos, a fim de compreender as variáveis.

A Análise Descritiva é a fase inicial do processo de estudo dos dados coletados. Utiliza-se métodos estatísticos descritivos para organizar, resumir e descrever aspectos importantes dos dados, um conjunto de características observadas e/ou uma comparação dessas características entre dois ou mais. Logo, ao realizar ou desenvolver um modelo, após a coleta das informações, faz-se necessário analisar as variáveis que serão utilizadas, podendo fazer uma análise univariada, bivariada (comparando a relação existente entre duas variáveis) ou até mesmo multivariada (que aborda a complexidade resultante da multiplicidade das variáveis).

Afinal, o que são variáveis, é um termo usado para descrever algo que pode ser medido e também pode variar. Variáveis aleatórias (v.a.) são funções que associam cada elemento de um conjunto a um número. Formalmente, são funções que atribuem elementos do espaço amostral a um conjunto de números reais. Existem dois tipos de variáveis aleatórias, as Variáveis Aleatórias Discretas e as Variáveis Aleatórias contínuas, como mostra a Figura 1.

Figura 1: Variável Aleatória



Fonte: Autoras, 2022.

A variável discreta é quantitativa, pois é utilizada para figurar as quantidades de um item específico, por exemplo: o número de reclamações de uma empresa, o número de defeitos em um produto ou serviço, o número de copos de água ingeridos por dia, a intento. Assim, segue-se que as discretas são sempre numéricas.

O número de carros que uma pessoa possui é uma variável discreta. Por exemplo, um homem pode ter um carro, dois ou três carros, só para citar algumas possibilidades, mas você não pode ter 1,6 carros ou 2,8 carros. Do mesmo modo o número de filhos que uma mulher tem, também é uma variável discreta, pode-se ter 2, 4 ou 6 filhos, nunca 2,1 ou 5,78 filhos.

Por outro lado, variáveis contínuas que podem assumir qualquer valor em um intervalo, e sempre existem outros valores intermediários entre dois valores observáveis. Ter mais ou menos valores depende da precisão da metrologia. Por exemplo, a altura de uma criança pode ser de 1,2 metros, 1,24 metros ou 1.249 metros, dependendo de como são medidas. Uma variável é chamada de contínua se pode assumir qualquer valor entre seus valores mínimo e máximo; caso contrário, é chamada de discreta.

Variáveis contínuas podem assumir qualquer valor dentro de um intervalo, também são consideradas métricas ou quantitativas, onde pode ter um número ou valor infinito entre dois pontos dados. Normalmente, medem-se as contínuas em escala. Por exemplo, quando se mede altura ou peso ou temperatura, tem-se dados contínuos.

Em experimentos científicos, são usadas como forma de agrupar os dados. Normalmente, uma variável é uma característica de um conjunto de objetos ou eventos que podem ser medidos com muitos valores numéricos diferentes. Como mostrado acima, é impossível ter quantidade de filhos não inteiros em uma família. No entanto, se você medir a altura, a variável será contínua, pois há um número infinito de possibilidades, mesmo olhando apenas entre 1 e 1,1 metros.

Tanto as variáveis discretas quanto as contínuas podem assumir um papel importante em um experimento científico. Durante um experimento, o cientista geralmente deseja observar os resultados dessa mudança. Muitas vezes, apenas uma é alterada porque determinaria a causa da resposta relevante se as variáveis fossem influenciadas.

Por exemplo, durante um experimento, a quantidade de luz que ilumina uma planta pode mudar, a quantidade de luz seria a variável independente. Para medições repetíveis, a variável independente é provavelmente uma variável discreta, como uma hora, duas horas ou horas-luz. A resposta da planta, seu grau de crescimento, a direção em que ela cresce, será a variável dependente contínua, dado que a quantidade de crescimento da planta possui um número infinito de resultados.

## 2.1 MEDIDAS DE TENDÊNCIA CENTRAL

Em geral, medidas de tendência central são aquelas que indicam um ponto em torno de dados. Isso nada mais é do que um centro de distribuição de dados. Para melhor compreensão, segundo Reis (1998):

*“O valor a escolher depende das características dos dados. Por exemplo, num estudo agrícola sobre a produção de trigo por hectare de terra arável podemos estar interessados em conhecer o valor mais elevado da produtividade do solo agrícola das várias explorações analisadas. Num outro estudo sobre os resultados de uma turma de estudantes universitários talvez seja mais interessante conhecer o resultado médio obtido por 50% dos estudantes. Num outro estudo sobre os rendimentos per capita dos países da CEE, a comparação entre países será facilitada se calcularmos os rendimentos médios de cada país.” (REIS, 1998, p.54. )*

As medidas de tendência central mais utilizadas são: média aritmética, média ponderada, moda e mediana. Essas medições podem fornecer um único valor que representa toda a quantidade de dados.

### 2.1.1 Média Aritmética

A média aritmética ou média de um conjunto de  $N$ , com os valores

$X = \{x_1, x_2, x_3, \dots, x_n\}$  usualmente representado por  $\bar{X}$ , é definida por:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \quad 1 \leq i \leq n$$

Quando os valores  $x_1, x_2, x_3, \dots, x_n$  têm associados a eles certos fatores de peso, ou ponderação,  $p_1, p_2, p_3, \dots, p_n$ , que os distinguem em importância relativa dentro de um conjunto de valores, a média ponderada é definida por:

$$\bar{X} = \frac{\sum_{i=1}^n P_i X_i}{\sum P_i}, \quad 1 \leq i \leq n$$

Para se obter a média ponderada, calcula-se o produto de cada valor por seu respectivo peso, soma-se esses produtos e em seguida divide-se pela soma dos pesos.

## 2.1.3 Moda

Em um conjunto de dados, a moda é o resultado ou o intervalo que ocorre com maior frequência no conjunto, ou seja, o resultado com a maior frequência absoluta. A moda pode não existir (quando todos ocorrem com a mesma frequência), mesmo que exista, pode não ser única (quando há mais de um elemento com frequência máxima)

**Exemplo 1.1:** Conjunto no qual se tem apenas uma moda

No conjunto de números {2, 2, 3, 5, 5, 5, 8, 9} a moda é 5. – Variável Discreta

Em uma sala foi feito uma pesquisa sobre a altura dos alunos, o conjunto das alturas foi o seguinte {1,60; 1,55; 1,63; 1,63; 1,55; 1,65; 1,63; 1,58} . A moda nesse caso, é de 1,63 – Variável Contínua

**Exemplo 1.2:** Conjunto amodal, não possui nenhum elemento se repetindo.

O conjunto {2, 3, 5, 7, 15, 8, 9} , conjunto amodal – Variável discreta

O conjunto amodal de variável contínua – {150,50; 35,60; 56,70; 115,90}

**Exemplo 1.3:** Conjunto bimodal, possui duas modas, ou seja, dois elementos que se repetem.

No conjunto de números {1, 2, 2, 5, 7, 7, 3} as modas são 2 e 7 – variável discreta

No conjunto dos números {3,5; 7,6; 7,6; 7,6; 8,5; 8,5; 8,5; 4,5; 9,2} as modas são 7,6 e 8,5 – variável contínua.

Em um histograma de frequência, a moda será sempre os valores que ocorrem com maior frequência. Distribuições com um único pico (valor máximo) são ditas unimodais.

#### 2.1.4 Mediana

A mediana de um conjunto de números ordenados é o valor central que divide o conjunto em, aproximadamente, 50% dos valores abaixo e 50% acima dele. Na prática, para determinar esse valor, primeiramente é necessário colocar os elementos em rol, ou seja, organizar os elementos do conjunto em ordem crescente ou decrescente, após ter realizado isso observa-se que, quando o número de elementos for ímpar, a mediana será o elemento do meio da sequência ordenada. Quando o número de elementos for par, a mediana será a média aritmética dos dois valores centrais.

**Exemplo 1.4:** Quando o conjunto tem uma quantidade ímpar.

No conjunto de números {1, 2, 3, 4, 5, 6, 7} a mediana é 4 – variável discreta

No conjunto dos números {4,5; 5,1; 5,3; 6,9; 7,1} a mediana é 5,3 – variável contínua.

**Exemplo 1.5:** Quando o conjunto tem uma quantidade par.

No conjunto de números {1, 2, 3, 4, 6, 6, 7, 8} a mediana é 5, pois os números centrais são 4 e 6, desta forma,  $\frac{4+6}{2} = \frac{10}{2} = 5$

No conjunto de números {2,4; 2,8; 3,8; 4,5; 4,9; 5,4} a mediana é 4,15, pois os números centrais são 3,8 e 4,5, desta forma,  $\frac{3,8+4,5}{2} = \frac{8,3}{2} = 4,15$

As medidas de dispersão são utilizadas para conhecer qual o grau de variação dos dados, pois afinal, nenhuma medida é, de fato, exata, já que há diversos fatores que podem influenciar a coleta de dados, como o instrumento utilizado ou a amostra que estamos estudando.

São parâmetros estatísticos usados para determinar o grau de variabilidade dos dados de um conjunto de valores. A utilização desses parâmetros torna a análise de uma amostra mais confiável, visto que as variáveis de tendência central (média, mediana, moda) muitas vezes escondem a homogeneidade ou não dos dados.

Por exemplo, considera que um DJ de festas infantis selecione as músicas de acordo com a média das idades das crianças convidadas para uma festa, em que as idades de dois grupos de crianças que irão participar de duas festas diferentes são descritas a seguir:

Festa A: 1 ano, 2 anos, 4 anos, 12 anos e 13 anos

Festa B: 5 anos, 6 anos, 7 anos, 8 anos e 9 anos

Em ambos os casos, a média é igual a 7 anos de idade. Entretanto, ao observar as idades dos participantes pode-se admitir que as músicas escolhidas são iguais? Neste exemplo, a média não é uma medida eficiente, pois não indica o grau de dispersão dos dados. As medidas de dispersão mais usadas são: amplitude, variância, desvio padrão e coeficiente de variação.

### 2.2.1 Amplitude

A amplitude de um conjunto, em Estatística, é a diferença entre o maior e o menor valor da variável desse conjunto. Em outras palavras, para encontrar a amplitude de uma lista de números, basta subtrair o menor elemento do maior. Isto é,  $A = X_{\text{maior}} - X_{\text{menor}}$ . Por ser uma medida que não leva em consideração como os dados estão efetivamente distribuídos, não é muito utilizada.

"Suponha que dois alunos tenham alcançado a mesma média na escola: 7,0. As notas do primeiro aluno foram: 8,0; 7,0; 7,0 e 6,0. Já as notas do segundo foram 4,0; 5,0; 9,0 e 10,0".

No exemplo dado acima, existem duas amplitudes a serem avaliadas:

a do primeiro e a do segundo aluno. O primeiro aluno tem 8 como maior nota e 6 como menor. A amplitude de suas notas foi:  $8 - 6 = 2$ . O segundo aluno teve 10 como sua maior nota e 4 como menor. A amplitude de suas notas foi  $10 - 4 = 6$ . Embora não seja possível determinar qual dos dois teve um melhor desempenho apenas por essa medida – pois não é possível saber qual dos dois teve um aumento nas notas –, esses resultados já dizem que a variação de notas do primeiro aluno foi muito menor do que a do segundo.

### 2.2.2 Variância

A variância é uma das medidas que leva em consideração a totalidade dos valores da variável em estudo, e não apenas os valores externos, como a amplitude total. Por isso, essas medidas são índices de variabilidade bastantes estáveis e, conseqüentemente, muito utilizados no cotidiano. Além disso, a variância complementa as informações obtidas pelas medidas de tendência central, é uma das medidas de variabilidade mais importantes.

A variância amostral é dada pelas somas dos desvios ao quadrado dos valores observados e da média amostral, dividido pelo tamanho da amostra menos 1. Assim como na equação abaixo:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Em que:

$x_i$ : i-ésimo elemento

$\bar{x}$ : média amostral

n: tamanho da amostra

Em outras palavras, a distância entre  $x_i$  e  $\bar{x}$  auxilia no entendimento

quando os valores observados estão próximos ou não uns dos outros.

O desvio padrão, é a raiz quadrada da variância. O objetivo de sua utilização é voltar à unidade original dos dados, que no cálculo da variância ficaram elevados ao quadrado.

$$S = \sqrt{S^2}$$

Em que  $S^2$  representa a variância.

#### 2.2.4 Coeficiente de variação

O coeficiente de variação é uma medida relativa de variabilidade, útil para comparação entre conjuntos de dados, pois quantifica a concentração dos dados em torno da média de séries distintas. O coeficiente de variação é dado por:

$$CV = \frac{s}{\bar{X}} \times 100\%$$

Em que,  $s$  representa o desvio padrão amostral e  $\bar{X}$  representa a média da amostra.

### 2.3 SITUAÇÃO PROBLEMA

No presente capítulo, vimos alguns conceitos da estatística descritiva, estas que são importantes para se fazer uma análise univariada e bivariada dos dados, analisar o comportamento e o tipo de cada variável. A seguir apresentamos um exemplo, onde será feito um estudo dos dados.

Os dados da tabela a seguir são referentes a uma pesquisa realizada pela prefeitura de uma determinada cidade, que elaborou questões a fim de recolher informações dos moradores em relação a frequência no parque principal da cidade. Numa manhã de terça-feira, 20 pessoas foram entrevistadas e cada uma respondeu a questões para identificar sexo, idade, número de vezes que frequenta o parque por

semana, estado civil, meio de transporte utilizado para chegar ao parque e tempo de permanência nele.

Tabela 1 – Dados da situação problema

SEXO	IDADE	FREQUÊNCIA POR SEMANA	ESTADO CIVIL	MEIO DE TRANSPORTE	TEMPO DE PERMANÊNCIA
FEMININO	26	2	CASADO	A PÉ	30m
FEMININO	41	1	CASADO	A PÉ	40m
MASCULINO	49	3	CASADO	CARRO	1h
FEMININO	19	1	SOLTEIRO	MOTO	1h30m
MASCULINO	20	1	SOLTEIRO	A PÉ	40m
FEMININO	27	2	SOLTEIRO	CARRO	40m
MASCULINO	38	2	CASADO	CARRO	1h
MASCULINO	26	3	CASADO	CARRO	20m
FEMININO	16	1	SOLTEIRO	A PÉ	1h
MASCULINO	42	1	CASADO	A PÉ	1h
FEMININO	62	3	VIÚVO	MOTO	2h
FEMININO	21	3	CASADO	CARRO	1h
MASCULINO	32	2	CASADO	CARRO	40m

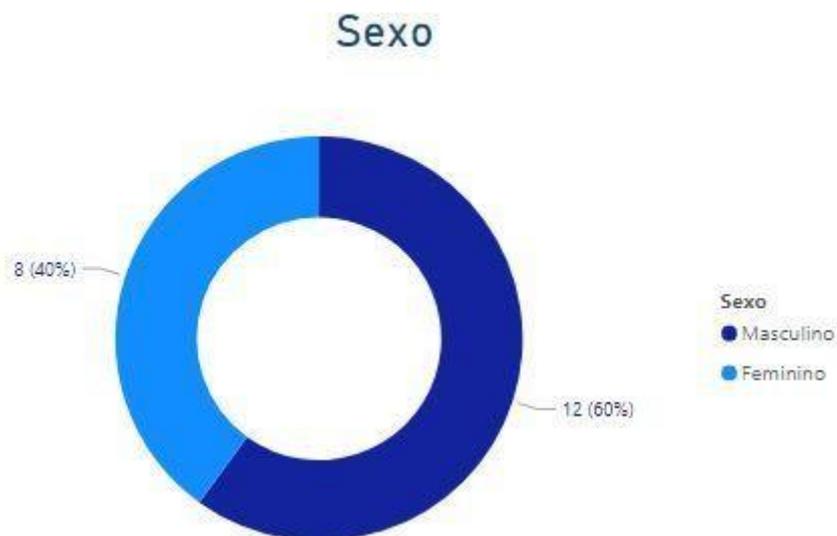
Fonte: Autoras, 2022

Assim, temos seis variáveis, onde podemos classificar como:

- Qualitativa Nominal: Sexo, Estado Civil, Meio de transporte;
- Quantitativa discreta: idade, frequência semanal no parque;
- Quantitativa contínua: tempo de permanência no parque.

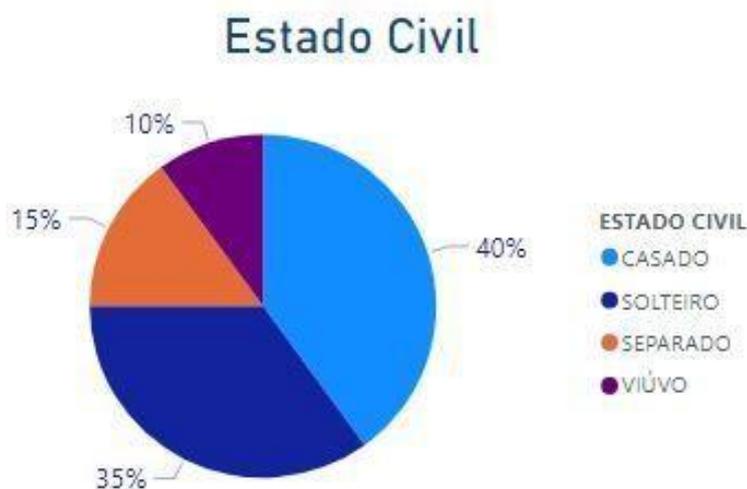
É importante analisar o tipo de variável e assim escolher o gráfico ideal e as métricas eficientes. A seguir, é apresentado dois gráficos para representar as variáveis qualitativas, que é o gráfico de rosca para a variável Sexo, conforme figura 2, este que mostra que o maior público é do sexo masculino; e o gráfico de pizza conforme figura 3 para a variável estado civil, onde podemos visualizar que está bem equilibrado entre solteiros ou casados.

Figura 2: Variável - Sexo



Fonte: Autoras, 2022

Figura 3: Variável - Estado Civil



Fonte: Autoras, 2022

Para a variável frequência semanal no parque, foi utilizado um gráfico de barras, conforme a figura 4, onde se pode identificar que a moda e a medianados dados são três.

Figura 4: Variável - Frequência Semanal



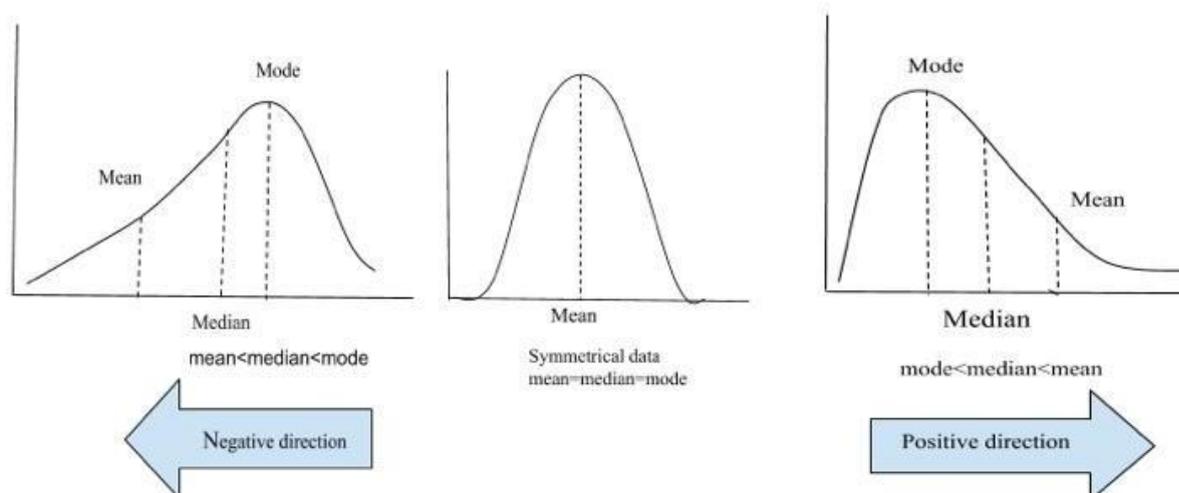
Fonte: Autoras, 2022

Além disso, temos que a média da frequência também é três.

$$\bar{X} = \frac{3 \times 1 + 4 \times 2 + 7 \times 3 + 3 \times 4 + 1 \times 5 + 1 \times 6 + 1 \times 7}{20} = 3$$

Quando as três medidas de centralidade são coincidentes, temos uma distribuição dos dados **simétrica**, e estas medidas são adequadas para representar os dados. Porém, pode haver outras situações, quando a moda é maior do que a mediana, e a mediana é maior do que a média tem-se uma distribuição assimétrica à esquerda, ou de **assimetria negativa**. E, quando a média é maior do que a mediana, e a mediana é maior do que a moda tem-se uma distribuição assimétrica à direita, ou de **assimetria positiva**. As três situações serão representadas no gráfico a seguir.

Figura 5: Simétrica, assimetria negativa, assimetria positiva



Fonte: Sanja Dados, 2021.

Entretanto, há situações em que as medidas de centralidade não são uma boa representação dos dados, como é o caso da variável idade.

- Média aritmética

$$\bar{X} = \frac{19+20+23+26+2 \times 27+28+29+2 \times 31+ \dots +56}{20} = \frac{716}{20} = 35,8$$

- Moda: 27, 31, 41
- Mediana: 33,50

Sendo assim, será calculado as medidas de dispersão, pois se tem um grau de confiabilidade maior.

- Variância:

$$s^2 = \frac{282,24 + 249,64 + 163,84 + 96,04 + 77,44 + \dots + 408,04}{20} = \frac{2421,2}{20} = 121,06$$

- Desvio Padrão:

$$s = \sqrt{s^2} = \sqrt{121,06} \approx 11,29$$

Desta forma, pode-se notar que há uma variação entre as idades.

### 3 TEORIA DA PROBABILIDADE

Ao se falar do conceito de métodos estatísticos, já se pensa em algo incerto, que se terá vários resultados possíveis, ou seja, pode-se pensar na palavra aleatório. Desse conceito, pensa-se sobre experimento aleatório, que de acordo com Morgado et al. (2001) é um experimento que repetido nas mesmas condições inúmeras vezes, não fornece impreterivelmente os mesmos resultados. Essa concepção diz então que mesmo que se conheça todas as variáveis e se tenha domínio sobre elas, o desfecho poderá não ser o mesmo, embora seja repetido perante circunstâncias idênticas. É possível relacionar um conjunto com todas as probabilidades de resultado do experimento, esse conjunto se chama Espaço amostral ( $\Omega$ ). Ao efetuar um grande número desse experimento, uma regularidade pode aparecer.

Em todos os experimentos aleatórios, tem-se o Espaço Amostral ( $\Omega$ ) que é o conjunto de todos os possíveis resultados, sendo fundamental identificá-lo para a compreensão e cálculo das probabilidades de eventos específicos. Para a desenvoltura e observação do espaço amostral depende-se de dois tipos de raciocínio, o contrafactual que é situações que não ocorreram, mas que poderiam ter ocorrido, e o raciocínio combinatório.

O raciocínio combinatório pode ser definido como uma forma de pensar que existe na análise de situações em que dados alguns conjuntos, seus elementos devem ser agrupados de forma que determinados critérios (seleção e/ou ordenação de elementos) sejam atendidos e determinantes – diretos ou indiretos – o número total de agrupamentos possíveis (BORBA, 2010).

#### 3.1 CONJUNTOS

Para entender a probabilidade e seus modelos, é necessário ter uma compreensão da teoria dos conjuntos pois essa teoria é a base da probabilidade.

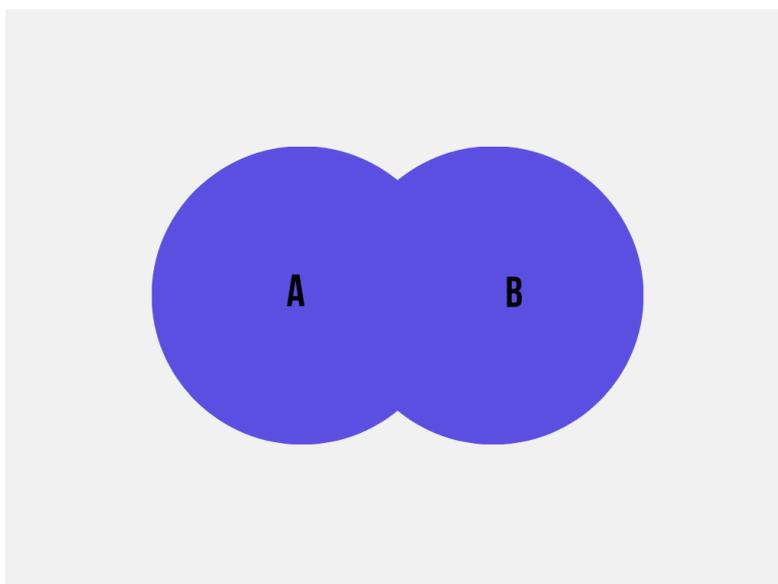
Um conjunto é uma coleção de objetos chamados membros ou elementos. Usualmente, os conjuntos são representados por letras maiúsculas, e as letras minúsculas representando os elementos do conjunto. Os elementos

pertencentes a um conjunto são escritos e lidos como pertencentes. O conjunto que contém todos os elementos aprendidos é chamado de conjunto básico, e o conjunto que não contém nenhum elemento é chamado de conjunto vazio ou nulo, denotado simbolicamente  $\emptyset$ .

Diz-se que  $A$  é subconjunto de  $B$  se todos os elementos de  $A$  forem elementos de  $B$  e escreve-se  $A \subset B$ . Dizemos que  $A = B$  se e somente se  $A \subset B$  e  $B \subset A$ , ou seja, são iguais se ambos contêm os mesmos elementos. A partir de conjuntos dados, pode-se construir novos conjuntos e/ou uni-los por meio de algumas operações básicas de matemática, tais como adição e multiplicação. Essa relação resultará em um novo conjunto, denotado por  $A \cup B$ , assim mostrado na figura 6.

$$A \cup B = \{x \in A \text{ ou } x \in B\}$$

Figura 6: Diagrama de Venn para  $A \cup B$ .

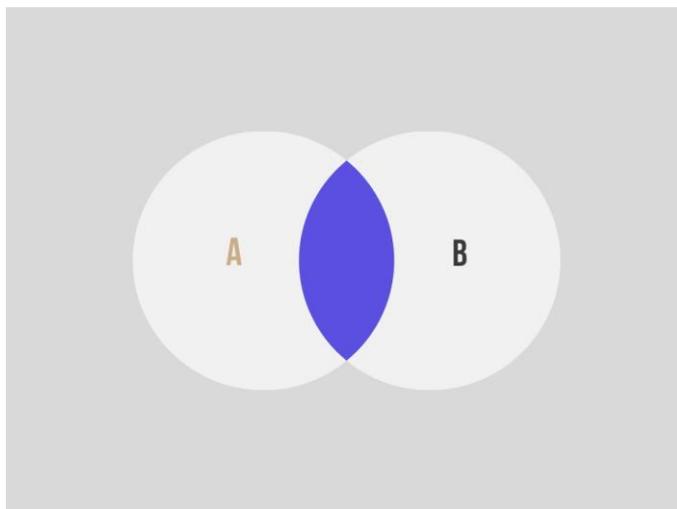


Fonte: Autoras, 2022

$A \cup B$  contém todos os elementos que estão em  $A$  e  $B$ .

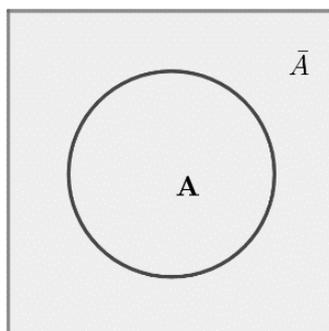
A intersecção de  $A$  e  $B$  é um conjunto cujos elementos estão tanto em  $A$  quanto em  $B$ . Denotaremos por  $A \cap B$  como o conjunto resultado da intersecção de  $A$  e  $B$ , conforme figura 3

$$A \cap B = \{x \in A \text{ e } x \in B\}$$

Figura 7: Diagrama de Venn para  $A \cap B$ 

Fonte: Autoras, 2022

Dado o conjunto fundamental  $U$ ,  $A \subset U$ , e o conjunto dos elementos de  $U$  que não estão em  $A$  é definido como conjunto complementar de  $A$ , denotado por  $\bar{A}$  de acordo com a figura 4, em que

Figura 8: Diagrama de Venn para  $\bar{A}$ .

Fonte: Autoras, 2022

### 3.2 ESPAÇO AMOSTRAL

Um experimento é dito aleatório se tiver mais de um resultado possível, caso tenha apenas um, é dito determinado. É qualquer experimento que pode ser

repetido infinitamente e tem um conjunto bem definido de resultados possíveis, conhecido como espaço amostral.

O espaço amostral denotado por  $\Omega$  é o conjunto de todos os resultados possíveis de um experimento aleatório. O conjunto pode ser exibido com mais ou menos detalhes dependendo do que se pode observar em um determinado problema. Considerando um experimento de jogar uma moeda duas vezes. Tem-seo lado “cara (C)” e o outro “coroa (K)”. O espaço amostral deste experimento é:

$$\Omega = \{CC, CK, KC, KK\}$$

### 3.3 EVENTOS

Os eventos relacionados ao experimento são um subconjunto dos possíveis resultados contidos em  $\Omega$ . Se um evento consiste em um único resultado, é chamado de evento único ou determinístico; se consiste em vários resultados possíveis, é chamado de evento composto. E o conjunto vazio é o evento impossível.

Dado dois eventos, a união entre eles, é o evento em que pelo menos um dos dois ocorre. A interseção entre os eventos, é o que ambos ocorrem.

**Exemplo 3.1:** Seja o espaço amostral o lançamento de um dado, logo temos

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

E os seguintes eventos,

$$A = \{\text{ocorrer número par}\} = \{2, 4, 6\}$$

$$B = \{\text{ocorrer número maior ou igual a 5}\} = \{5, 6\}$$

Logo, temos

- União:  $A \cup B = \{2, 4, 5, 6\}$
- Interseção:  $A \cap B = \{5\}$
- Diferença:  $A - B = \{2, 4\}$
- Complementar:  $\bar{A} = \{1, 3, 5\}$

A teoria das probabilidades atribui números aos eventos que representam a probabilidade de ocorrência desse evento. Dado um evento, esse número associado é chamado de probabilidade e é simbolizado por  $P$ . A definição original de probabilidade era o quociente da divisão do número de resultados possíveis de um evento pelo número total de resultados possíveis de um experimento.

Por exemplo, considerando o exemplo 3.1, temos

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{3}{6} = \frac{1}{2}$$

No entanto, esta definição só se aplica se todos os resultados possíveis tiverem a mesma probabilidade de ocorrer e o espaço amostral for finito. A definição dada mais tarde foi a de frequência relativa. Após repetir o experimento várias vezes nas mesmas condições, a frequência de ocorrência do evento é observada, e a frequência relativa é o número de ocorrências. À medida que aumenta, a frequência relativa tende a se estabilizar e atingir um limite. Esta frequência relativa é definida como a probabilidade.

Considere um experimento aleatório com espaço amostral  $\Omega$ , finito, isto é,  $\Omega = \{a_1, a_2, \dots, a_k\}$ . Suponha que o experimento seja repetido  $N$  vezes, nas mesmas condições. Seja  $n_i$  o número de vezes que ocorre o evento elementar  $a_i$ . Definimos **frequência relativa do evento**  $\{a_j\}$  como sendo o número  $f_j$ , tal que

$$f_i = \frac{n_i}{N} \quad \forall i \in \{1, 2, \dots, k\}$$

A cada **evento elementar**  $\{a_j\}$  associando um **número real**, indicado por  $p(\{a_j\})$  ou  $p_j$ , chamado **probabilidade do evento**  $\{a_j\}$  satisfazendo as seguintes condições:

1.  $0 \leq p_i \leq 1, \forall i \in \{1, 2, \dots, k\}$
2.  $\sum_{i=1}^k p_i = p_1 + p_2 + \dots + p_k = 1$

Dizemos que os números  $p_1, p_2, \dots, p_k$  definem uma **distribuição de**

**probabilidades** sobre  $\Omega$ .

Em seguida, seja  $A$  um evento qualquer de  $\Omega$ . Definimos **probabilidade do evento  $A$**  (e indicamos por  $P(A)$  da seguinte forma:

a) Se  $A = \emptyset, P(A) = 0$

b) Se  $A \neq \emptyset, P(A) = \sum_{a_i \in A} p_i$

Isto é, a probabilidade de um evento constituído por um certo número de elementos é a soma das probabilidades dos resultados individuais que constituem o evento  $A$ .

A seguir temos alguns teoremas sobre probabilidades em espaço amostral finito.

### 3.4.1 Teoremas da probabilidade em espaço amostral finito

**Teorema 1 :** “A probabilidade do evento certo é 1”

Demonstração: de fato, o evento certo é  $\Omega = \{a_1, a_2, \dots, a_k\}$  e por definição:

$$P(\Omega) = p_1 + p_2 + \dots + p_k = 1$$

**Teorema 2:** “Se  $A \subset B$ , então  $P(A) \leq P(B)$ ”

Demonstração: 1) Se  $A = B$ , por definição  $P(A) = P(B)$  e, entretanto,  $P(A) \leq P(B)$ .

2) Se  $A \neq B$ .

Seja  $A = \{a_1, a_2, \dots, a_r\}$  e  $B = \{a_1, a_2, \dots, a_r, a_{r+1}, \dots, a_{r+q}\}$ , então:

$$P(A) = p_1 + p_2 + \dots + p_r$$

$$P(B) = p_1 + p_2 + \dots + p_r + p_{r+1} + \dots + p_{r+q}.$$

$p_1 + p_2 + \dots + p_r + p_{r+1} + \dots + p_{r+q}$  são todos positivos, decorre que  

$$P(A) \leq P(B)$$

No caso particular de  $A = \emptyset$ , temos  $P(A) = 0$  e  $P(B) \geq 0$ , e portanto  

$$P(A) \leq P(B)$$

**Teorema 3:** “Se  $A$  é um evento, então  $0 \leq P(A) \leq 1$ . “

Demonstração:  $\emptyset \subset A \subset \Omega$ . Logo, pelo teorema 2:

$$P(\emptyset) \leq P(A) \leq P(\Omega) \text{ e então } 0 \leq P(A) \leq 1$$

**Teorema 4:** “Se  $A$  e  $B$  são eventos, então  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .”

Demonstração:  $P(A \cup B) = \sum_{a \in A \cup B} P_j$ . outro lado,

$$P(A) = \sum_{a \in A} P_j \text{ e } P(B) = \sum_{a \in B} P_j.$$

Quando somamos  $P(A) + P(B)$  as probabilidades dos eventos fundamentais contidos em  $A \cap B$  são contadas duas vezes, uma por estarem em  $A$  e outra, por estarem em  $B$ . No entanto,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  é a somatória das probabilidades dos eventos fundamentais contidos em  $A \cup B$ , logo:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

**Teorema 5:** “Se  $A$  é um evento, então  $P(\bar{A}) = 1 - P(A)$ ”

Demonstração: Como  $A \cap \bar{A} = \emptyset$  e  $A \cup \bar{A} = \Omega$ , decorre pelo teorema 4 que  $P(A \cup \bar{A}) = P(A) + P(\bar{A})$ , logo

$$1 = P(A) + P(\bar{A}) \Rightarrow P(\bar{A}) = 1 - P(A)$$

**Exemplo 3.2:** Uma urna contém 50 bolinhas numeradas, de 1 a 50. Uma bolinha é escolhida ao acaso e observado seu número. Admitindo

probabilidades iguais a  $\frac{1}{50}$  para todos os eventos fundamentais, qual a probabilidade

de:

- Observar um múltiplo de 3 e 4 ao mesmo tempo?
- Observar um múltiplo de 3 ou de 4?
- Observar um número não múltiplo de 5?

Temos:  $\Omega = \{1, 2, 3, \dots, 49, 50\}$

- Um múltiplo de 3 e 4 simultaneamente, tem que ser múltiplo de 12; então, o evento que nos interessa é  $A = \{12, 24, 36, 48\}$ .

$$P(A) = \frac{1}{50} + \frac{1}{50} + \frac{1}{50} + \frac{1}{50} = \frac{4}{50} = \frac{2}{25}$$

- Sejam os eventos:

B: número múltiplo de 3.

C: número múltiplo de 4.

O evento que irá interessar é  $B \cap C$ , então:

$B = \{3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48\}$ , logo

$$P(B) = \frac{16}{50} = \frac{8}{25}$$

$C = \{4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48\}$ . Logo,

$$P(C) = \frac{12}{50} = \frac{6}{25}$$

Portanto,  $P(B \cup C) = P(B) + P(C) - P(B \cap C)$ ,  $B \cap C$  nada mais é

do que o evento A, resolvido no item a, logo  $P(B \cap C) = \frac{2}{25}$ .

Segue-se então que

$$P(B \cup C) = \frac{8}{25} + \frac{6}{25} - \frac{2}{25} = \frac{8+6-2}{25} = \frac{12}{25}$$

- Seja D o evento, o número é múltiplo de 5. Temos:

$D = \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$

$$P(D) = \frac{10}{50} = \frac{1}{5}$$

O evento que interessa é  $\bar{D}$ , logo

$$P(\bar{D}) = 1 - P(D) = 1 - \frac{1}{5} = \frac{4}{5}$$

Enfim, ao se trabalhar com dados amostrais, primeiro é feita uma análise descritiva dos dados, em seguida uma exploração da teoria da probabilidade dentro de cada variável estudada e por fim técnicas de inferência para auxiliar na tomada de decisão sobre a população. Desta forma, no próximo capítulo, será

abordado alguns tipos de distribuição de probabilidades, utilizadas em variáveis discretas ou contínuas e assim auxiliar em uma deliberação.

## 4 DISTRIBUIÇÕES DE PROBABILIDADE

Quando a Estatística é aplicada na resolução de problemas, muitos se apresentam com as mesmas características, o que permite estabelecer uma teoria para determinação da solução. As distribuições de probabilidade são modelos matemáticos que relacionam um certo valor da variável em estudo com a sua probabilidade de ocorrência.

Existem dois tipos de distribuição de probabilidade, as distribuições discretas e as distribuições contínuas.

### 4.1 DISTRIBUIÇÕES DISCRETAS

Distribuições de probabilidade discretas recebem esse nome porque são características de variáveis aleatórias discretas, isto é, variáveis que podem assumir valores numéricos finitos ou infinitos.

Em outras palavras, sendo uma variável aleatória  $x$ , os valores possíveis de  $x$  podem ser uma sequência de valores como  $x_1, x_2, x_3, \dots$ . No caso finito, a sequência possui um valor final  $x_n$ , e no caso infinito a sequência continua

indefinidamente. Exemplos de variáveis aleatórias discretas: número de animais, número de pessoas, número de objetos, etc.

As principais distribuições de variáveis discretas são a Distribuição binomial e a distribuição de Poisson. Em seguida, será abordado algumas propriedades de cada uma delas.

#### 4.1.1 Distribuição Binomial

Em teoria das probabilidades e estatística, a distribuição binomial é a distribuição de probabilidade discreta do número de sucessos numa sequência tentativas tais que:

Cada tentativa tem exclusivamente como resultado duas possibilidades, sucesso ou fracasso (binomial, a que se chama de tentativa de

Bernoulli), e; cada tentativa é independente das demais, a probabilidade  $p$  de sucesso a cada tentativa permanece constante independente das demais, e; a variável de interesse, ou pretendida, é o número de sucessos  $k$  nas  $n$  tentativas.

Na prática, existem muitos experimentos que admitem apenas dois resultados, como por exemplo: uma peça é classificada como boa ou defeituosa; o resultado de um exame médico para detecção de uma doença é positivo ou negativo; um paciente submetido a um tratamento, durante um período de tempo fixo, cura-se ou não da doença; um entrevistado concorda ou não com a afirmação feita.

Quando um fator ou condição permanece constante, segundo Meyer 2010, pode-se medi-lo e obter resultados desde que o experimento seja estável sendo assim, pode-se aplicar o modelo binomial

**Definição 01:** A variável  $k$  tem distribuição *Binomial* com parâmetros  $n$ ,  $p$  se sua função de probabilidade é dada por

$$P(k) = \binom{n}{k} p^k q^{n-k}, x = 1, 2, 3, \dots, n$$

Onde,

- $\binom{n}{k}$  = binomial
- $p$  = probabilidade de sucesso
- $q = (1 - p)$  = probabilidade de fracasso
- $n$  = número de eventos estudados
- $k$  = número de eventos desejados que tenham sucesso

**Exemplo 4.1:** A probabilidade de que um homem de 45 anos sobreviva mais 20 anos é 0,6. De um grupo de 5 homens com 45 anos, qual a probabilidade de que exatamente 4 cheguem aos 65 anos?

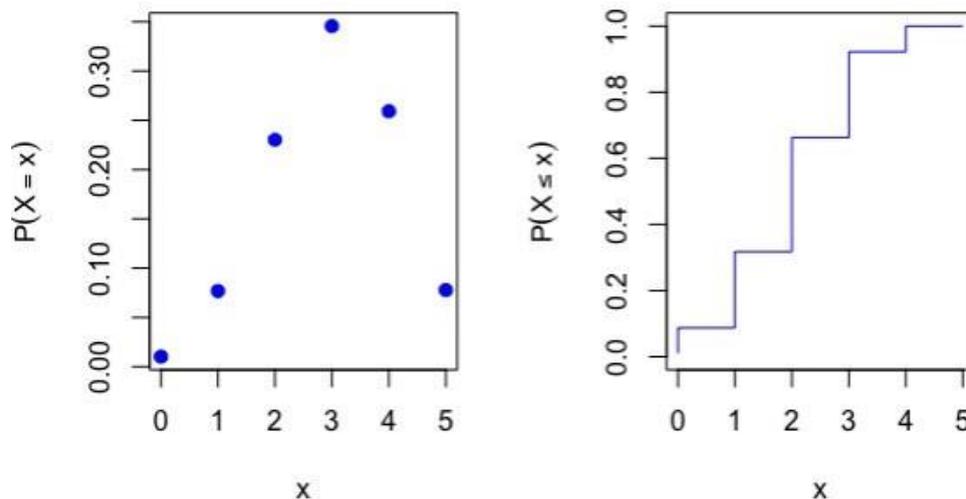
$$P(4) = \binom{5}{4} 0,6^4 \cdot 0,4^{5-4} = 0,2592$$

- Qual a probabilidade de que três sobrevivam?

$$P(0) + P(1) + P(2) + P(3) = 0,66304$$

A figura 9, representa os gráficos da distribuição binomial nos intervalos do exemplo anteriormente dado. Onde o primeiro gráfico apresenta a função de probabilidade para o Binomial de 0 a 5 com sucesso de 0,6. O segundo gráfico apresenta a distribuição acumulada para o binomial.

Figura 9: Gráficos distribuição binomial



Fonte: Autoras, 2022

#### 4.1.2 Distribuição de Poisson

As estatísticas de Poisson são usadas para modelar resultados aleatórios ao contar eventos com um intervalo médio definido. Várias formas de decaimento radioativo, número de bebês que nascem mensalmente em um certo hospital, fornecem exemplos disso.

Pode-se fazer a seguinte pergunta: se repetir um experimento muitas vezes, qual distribuição deve-se esperar que o número de contagens  $v$  seja observado em dois minutos? Diz-se que a distribuição esperada é uma distribuição binomial. Se houver  $n$  núcleos, e a probabilidade de qualquer núcleo decair é  $p$ , então a probabilidade de  $v$  decair é a probabilidade de  $v$  ter sucesso em  $n$  tentativas, ou  $B_{n,p}(v)$ . No entanto, neste tipo de experimento pode-se fazer uma simplificação importante. O número de tentativas, ou seja, o número de núcleos, é enorme.

Nesses casos, onde  $n$  é grande e  $p$  é pequeno, pode-se mostrar que a distribuição binomial é indistinguível de uma função mais simples chamada distribuição de Poisson.

**Definição 02:** A variável  $x$  tem distribuição *Poisson* de parâmetro  $\lambda$ ,  $\lambda > 0$ , se função de probabilidade é dada por

$$P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

onde:

- $e$  = número de Euler, cujo valor aproximado é 2,718281828459045
- $\lambda$  = representa o número médio de ocorrências em um determinado intervalo de tempo ou espaço.
- $x$  = número de sucessos no intervalo desejado.

**Exemplo 4.2:** Suponha que o número de pessoas utilizando um determinado caixa eletrônico em uma dada hora possa ser modelado por uma distribuição de Poisson com  $\mu = 5,5$ . Qual é a probabilidade de 8 pessoas a mais utilizarem a máquina durante a próxima hora?

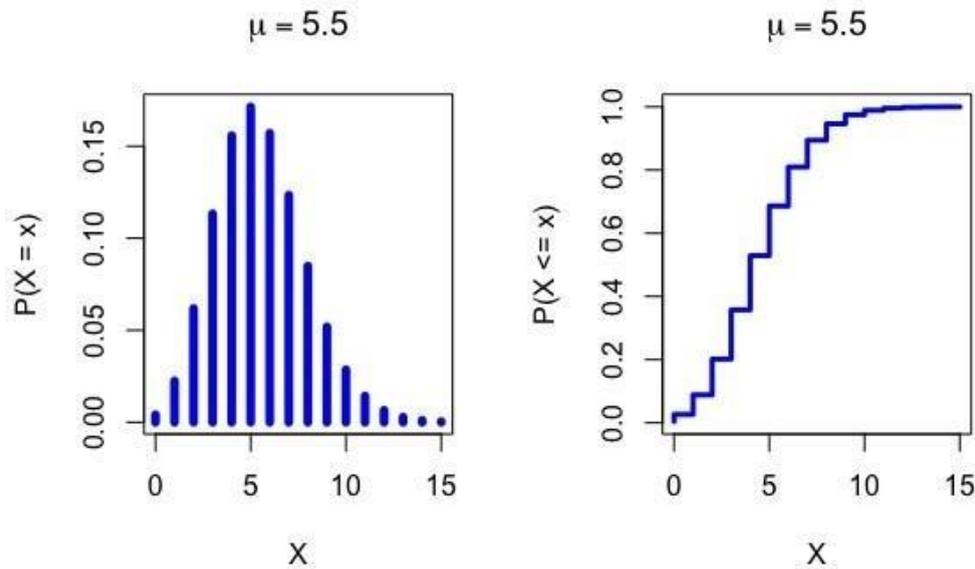
$$P(8) = \frac{e^{-5,5} 5,5^8}{8!} = 0,1056$$

Qual a probabilidade de até 8 pessoas utilizarem a máquina durante a próxima hora?

$$P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7) + P(8) = 0,8944$$

No primeiro gráfico abaixo, temos os resultados da função de probabilidade de Poisson e no segundo gráfico temos a distribuição acumulada de Poisson, considerando  $\mu = 5,5$ , com  $0 \leq x \leq 15$ .

Figura 10: Gráficos distribuição de Poisson



Fonte: Autoras, 2022.

## 4.2 DISTRIBUIÇÕES CONTÍNUAS

A distribuição contínua descreve as probabilidades dos possíveis valores de uma variável aleatória contínua, no qual é uma variável aleatória com um conjunto de valores possíveis que é infinito e incontável. As suas probabilidades são definidas como a área sob a curva da sua distribuição. Assim, apenas as faixas de valores podem ter uma probabilidade diferente de zero. A probabilidade de que uma variável aleatória contínua seja igual a algum valor é sempre zero.

Variáveis contínuas podem assumir praticamente qualquer valor numérico e podem ser significativamente divididas em incrementos menores. Normalmente, são medidas em escala. Por exemplo, quando se mede altura, peso e temperatura, se tem dados contínuos. Usando-as, é possível calcular e avaliar a média, mediana, desvio padrão ou variância.

Ela é contínua se for teoricamente possível para os membros do grupo cair em qualquer lugar em um espectro com pequenas quantidades de uma característica em uma extremidade e grandes quantidades de uma característica na outra extremidade, são frequentemente medidas em unidades infinitamente pequenas.

Em experimentos científicos, as variáveis são usadas como forma de agrupar os dados. Normalmente, é uma característica de um conjunto de objetos ou eventos que podem ser medidos com muitos valores numéricos diferentes. Como mostrado acima, é impossível ter tamanhos de filhos consecutivos em uma família. No entanto, se você medir a altura, a variável será contínua, pois há um número infinito de possibilidades, mesmo olhando apenas entre 1 e 1,1 metros.

É importante lembrar que esses dois tipos são agrupados de acordo com a escala em que são medidos e o que é medido. Na maioria dos experimentos científicos, escalas discretas são usadas para medir esses dois tipos de variáveis. Como existem infinitas possibilidades, isso significa que as medições das contínuas geralmente são arredondadas para facilitar o trabalho com os dados.

Tanto as variáveis discretas quanto as contínuas podem assumir um dos papéis em um experimento científico. Durante um experimento, o cientista geralmente deseja observar os resultados da mudança de uma variável. Muitas vezes, apenas uma variável é alterada porque determinaria a causa da resposta relevante se as variáveis fossem influenciadas.

Por exemplo, durante um experimento, a quantidade que ilumina uma planta muda. A quantidade de luz seria a variável independente. Para medições repetíveis, a qual é provavelmente uma variável discreta, como uma hora, duas horas ou horas-luz. A resposta de uma planta, seu grau de crescimento e a direção em que ela cresce, será a variável dado que a quantidade de crescimento da planta seja um número infinito de resultados, é um dependente contínuo.

#### 4.2.1 Distribuição Exponencial

A distribuição exponencial e de Poisson compartilham uma conexão estreita devido ao fato de que ambas lidam com a probabilidade de eventos ocorrerem ao longo do tempo. Essas duas distribuições podem ser usadas alternadamente para descrever as probabilidades envolvidas em vários eventos. A distribuição exponencial se refere ao tempo entre dois eventos e esse intervalo médio entre eles é equivalente à variável aleatória.

A distribuição exponencial permite medir o tempo e a distância entre os eventos da distribuição de Poisson. Ao analisar um jogo de futebol, consideramos o número de gols marcados, esse fenômeno é conhecido como variável aleatória de

Poisson, que exibe uma taxa de variação exponencial e pode-se estudar o intervalo entre as ocorrências de gols, e isso também se torna exponencial.

A distribuição exponencial é comumente usada para modelar falhas de componentes, bem como os tempos de espera em sistemas de filas.

**Definição 03** A variável  $x$  tem distribuição exponencial de parâmetro  $\lambda$ ,  $\lambda > 0$ , se tiver a densidade dada por

$$P(x) = \lambda e^{-\lambda x}$$

onde:

- $e$  Constante de Euler, cujo valor aproximado é de 2,718281828459045
- $\lambda$  Taxa (rate) de ocorrência dos eventos

**Exemplo 4.3:** Suponha que um determinado equipamento vital para a operação de seu negócio funcione ininterruptamente sem falhas, em média, por três meses. Qual é a probabilidade de haver uma avaria que implique no desligamento do equipamento em dois ou menos meses?

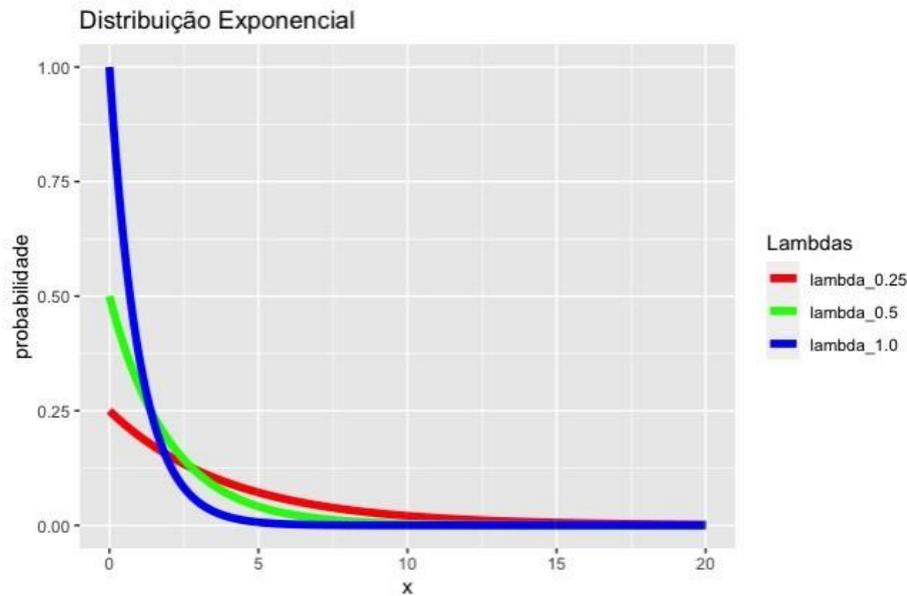
A taxa de avarias do equipamento é igual ao inverso do tempo médio entre falhas, conhecido como MTBF (Mean Time Between Failures). Portanto,

$$\lambda = \frac{1}{3}.$$

$$P(2) = \frac{1}{3} e^{-\frac{1}{3} \times 2} = 0,4865829$$

O gráfico abaixo mostra a função de distribuição exponencial, considerando diferentes valores para  $\lambda$ .

Figura 11: Gráfico Distribuição Exponencial



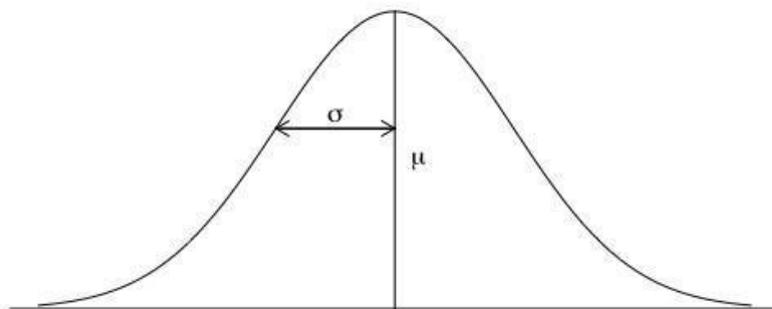
Fonte: Autoras, 2022

#### 4.2.2 Distribuição Normal

Distribuição normal, também chamado de modelo de Gauss, serve para explicar inúmeros fenômenos naturais, físicos, psicológicos, sociológicos, é extremamente importante em Estatística pois serve de fundamento para muitas técnicas de inferência e aproximações.

A distribuição normal é, talvez, a mais importante das distribuições de probabilidade. Erros de mensuração de fenômenos físicos ou econômicos são frequentemente modelados pela distribuição normal, mas esta não é a única aplicação desta densidade. Por exemplo, a distribuição dos pesos, alturas e QI's das pessoas numa população também já foram modelados com sucesso por esta distribuição. A distribuição normal tem a forma de um sino e possui dois parâmetros,  $\mu$  e  $\sigma$ , a sua média e desvio padrão, respectivamente.

Figura 12: Curva Normal



Fonte: Autoras, 2022

A distribuição normal é também chamada de gaussiana em homenagem ao matemático Carl Friederich Gauss (1777-1855), que a utilizou pela primeira vez na modelagem de erros de medida. A distribuição normal também funciona como uma boa aproximação para outras densidades. Por exemplo, sob algumas condições pode-se provar que a densidade binomial pode ser aproximada pela normal.

**Definição 04:** Dizemos que uma variável aleatória contínua  $x$  tem distribuição normal com parâmetros  $\mu$  e  $\sigma$ , se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ para } x \in \mathbb{R}$$

Em que:

$$\pi = 3,14159$$

$$e = 2,71828$$

$$\mu = \text{média}, -\infty \leq \mu \leq \infty$$

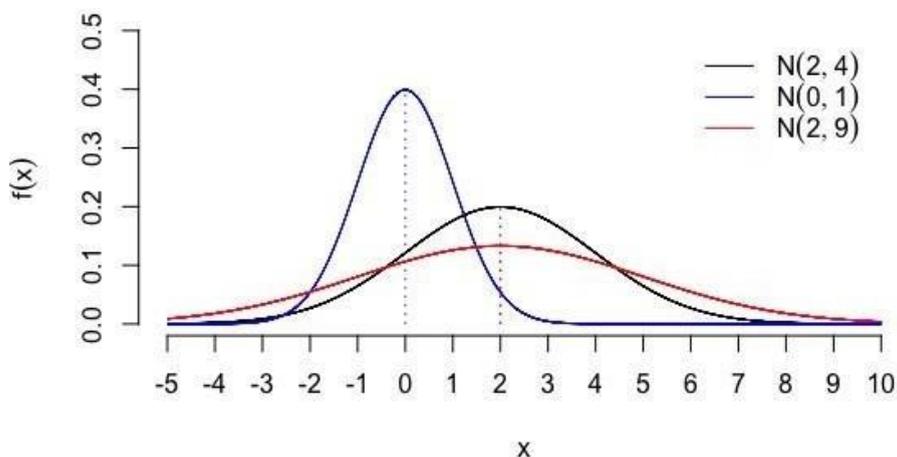
$$\sigma = \text{desvio padrão}, \sigma > 0$$

A distribuição normal possui as seguintes propriedades:

1. O ponto sobre o eixo horizontal, onde a curva tem um valor máximo, ocorre em  $x = \mu$ .
2. A curva é simétrica em torno de um eixo vertical que passa pela média  $\mu$ .
3. A curva tem seus pontos de inflexão em  $x = \mu \pm \sigma$ ; é côncava para baixo se  $\mu - \sigma$ .
4. A área compreendida entre a curva até o eixo das abscissas é igual a 1.
5.  $f(x) \rightarrow 0$ , então  $x \rightarrow \pm \infty$

O formato da Curva Normal dependerá dos parâmetros  $\mu$  (média) e  $\sigma$  (desvio padrão). Ao variar esses dois parâmetros temos movimentos de translação e achatamento no gráfico.

Figura 13: Curvas Normal com desvios e médias diferentes



Fonte: Autoras, 2022

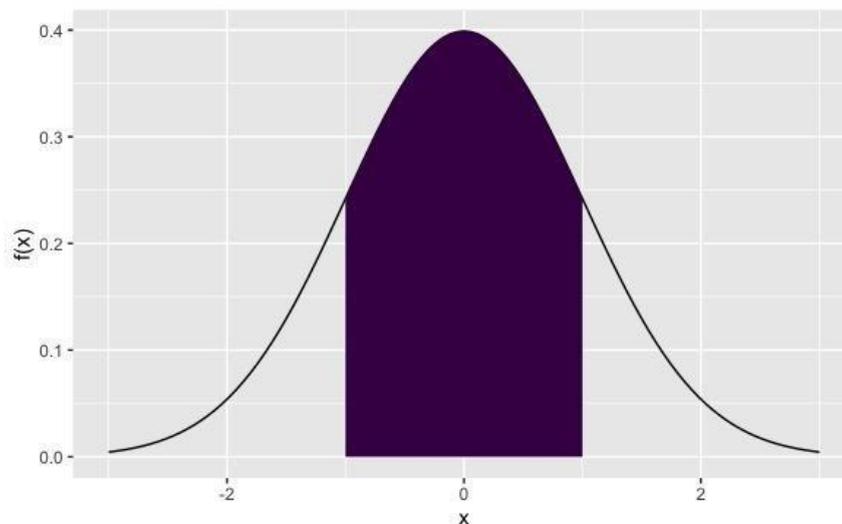
Podemos observar pelas representações gráficas que o valor da média indica o centro da distribuição e o desvio padrão mede a dispersão do conjunto, indicando a variabilidade em relação ao centro, ou seja, quanto maior o desvio padrão mais achatado será o gráfico.

Para calcular probabilidades com o Modelo Normal, devemos resolver a integral da função densidade de probabilidade no intervalo de interesse. Para uma distribuição contínua, portanto, não faz diferença se o intervalo de que estamos falando é aberto ou fechado. A probabilidade de que  $a \leq x \leq b$  é igual à probabilidade de que  $a < x < b$ , isto é,

$$P(a < x < b) = P(a \leq x \leq b) = \int_a^b f(x)dx$$

Assim a probabilidade de que um valor da variável aleatória de densidade de probabilidade  $f(x)$  seja observado no intervalo  $[a, b]$  é obtida pela área sob a curva  $f$  e o eixo das abscissas, em relação a esse intervalo, conforme mostrado na figura 14.

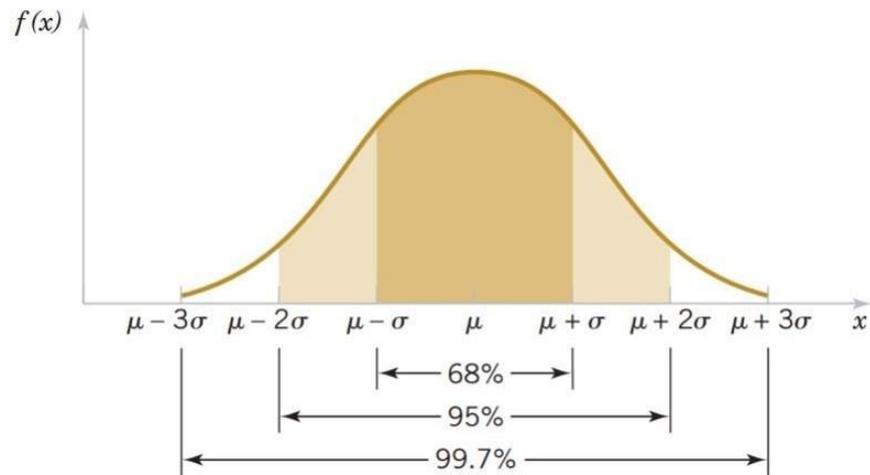
Figura 14: Área sob a curva Normal entre os pontos a e b



Fonte: Autoras, 2022.

No Modelo Normal, aproximadamente 68% da área total da curva está localizada entre  $\mu \pm \sigma$ , aproximadamente 95% da área total está localizada entre  $\mu \pm 2\sigma$  e aproximadamente 99,7% da área total da curva está localizada entre  $\mu \pm 3\sigma$ . Podemos observar esses fatos no gráfico da distribuição de frequência abaixo.

Figura 15: Distribuição da Curva Normal



Fonte: Probabilidade UFSC, 2016.

Para obter os valores numéricos correspondentes a estes fatos,

integramos a expressão  $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$  entre os limites apropriados:

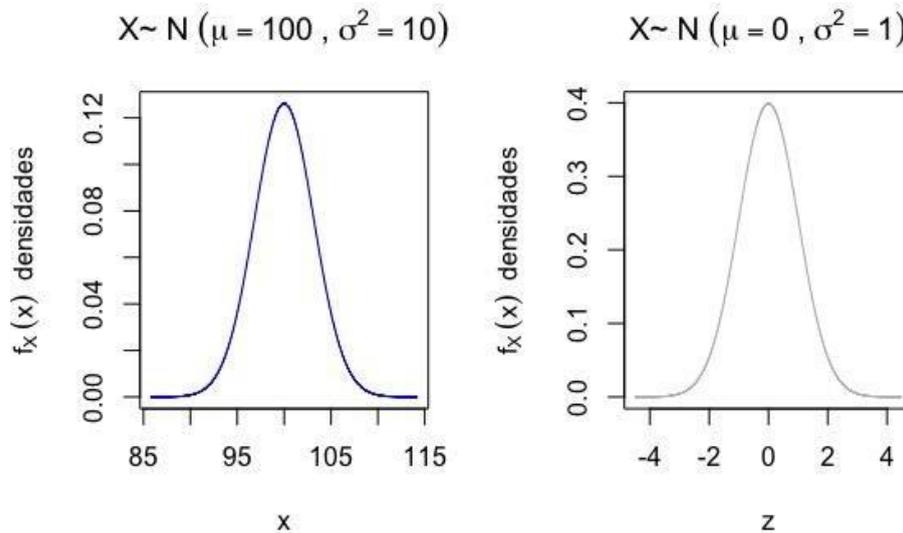
$$P(\mu - \sigma < x < \mu + \sigma) = \int_{\mu - \sigma}^{\mu + \sigma} f(x) dx = 0,6826 = 68,26\%$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = \int_{\mu - 3\sigma}^{\mu + 3\sigma} f(x) dx = 0,9973 = 99,73\%$$

Na prática, não é necessário calcularmos as integrais, podemos utilizar valores tabelados que permitem encontrar facilmente os valores desejados.

Porém é necessário padronizar a variável aleatória  $x$ , onde utilizamos a média zero e desvio padrão um, esses valores podem ser usados para fazermos inferências a respeito de qualquer Distribuição Normal. Na figura 14, podemos visualizar a mesma curva, porém, na primeira temos  $\mu = 100$  e  $\sigma^2 = 10$  e no segundo gráfico temos a curva com valores padronizados.

Figura 16: Curva normal  $X \sim N(100, 10)$  e  $X \sim N(0, 1)$



Fonte: Autoras, 2022

#### 4.2.2.2 Padronização de uma variável aleatória

Padronizar uma variável aleatória  $x$  de média  $\mu$  e desvio padrão  $\sigma$  consiste em construir a partir dela uma nova variável aleatória  $z$ , cujos valores são obtidos subtraindo-se de cada valor de  $x$  a média populacional e dividindo-se o resultado pelo desvio padrão

$$z = \frac{x - \mu}{\sigma}$$

Assim, para determinar a probabilidade de  $x \in [a, b]$ , temos

$$\begin{aligned} P(a \leq x \leq b) &= P(a - \mu \leq x - \mu < b + \mu) = \\ &= P\left(\frac{a - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \leq \frac{b + \mu}{\sigma}\right) = \\ &= P\left(\frac{a - \mu}{\sigma} \leq z \leq \frac{b + \mu}{\sigma}\right) \end{aligned}$$

Portanto, quaisquer que sejam os valores de  $\mu$  e  $\sigma$ , utilizamos a normal padronizada para obter probabilidades com a Distribuição Normal. Os valores da probabilidade são apresentados de acordo com o exposto no Anexo A.

**Exemplo 4.4:** O Inmetro verificou que as lâmpadas incandescentes da fabricante XPTO apresentam uma vida útil normalmente distribuída, com média igual a 720 dias e desvio padrão igual a 30 dias. Calcule a probabilidade de uma lâmpada escolhida ao acaso durar:

a. Menos que 700 dias?

$$Z = \frac{700-720}{30} = -\frac{2}{3} = -0,\bar{6}$$

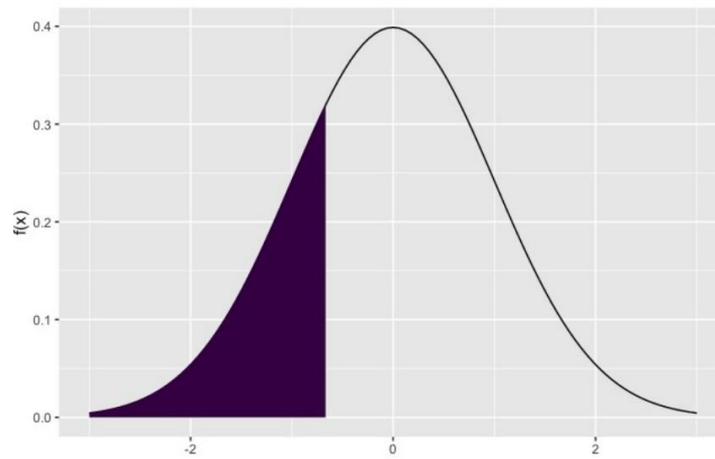
Consultando a tabela (Anexo A), conforme imagem abaixo:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365

Temos

$$P(700) = 1 - 0,7454 = 0,2546$$

Figura 17: Gráfico distribuição normal menos de 700 dias



Fonte: Autoras, 2022

b. Entre 650 e 750 dias?

$$Z_{inferior} = \frac{650-720}{30} = -\frac{7}{3} = -0,\bar{7} e$$

$$Z_{superior} = \frac{750-720}{30} = \frac{30}{30} = 1$$

Consultando a tabela no Anexo A, conforme imagem abaixo:

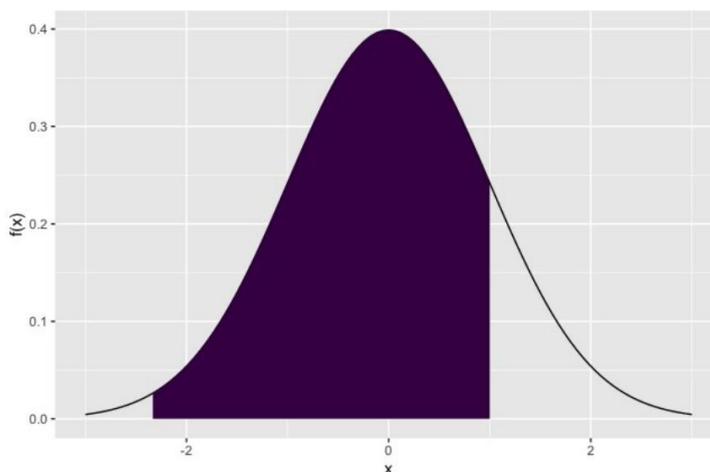
z	.00	.01	.02	.03	.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	0.9251
2.0	0.9772	0.9778	0.9783	0.9788	0.9793
2.1	0.9821	0.9826	0.9830	0.9834	0.9838
2.2	0.9861	0.9864	0.9868	0.9871	0.9875
2.3	0.9893	0.9896	0.9898	0.9901	0.9904

Temos

$$P(650 \leq x \leq 750) = P(750) - P(650) = 0,8413 - (1 - 0,9901)$$

$$P(650 \leq x \leq 750) = 0,8413 - 0,0098 = 0,8315$$

Figura 18: Gráfico distribuição normal entre 650 e 750 dias



Fonte: Autoras, 2022

c. Mais de 800 dias?

$$Z = \frac{800 - 720}{30} = \frac{80}{30} = 2,6\bar{6}$$

Logo, temos

$$P(x \geq 800) = 1 - 0,9962 = 0,0038$$

A distribuição normal, pode ser aplicada em diferentes situações, sendo uma das mais utilizadas, principalmente pelo fato da maioria dos dados apresentarem pelo menos trinta observações, sendo garantido pelo teorema do limite central, a distribuição normal.

Diversos modelos estatísticos aplicados à economia, negócios, saúde, utilizam a distribuição normal para prever vendas, melhor horário, dentre outros.

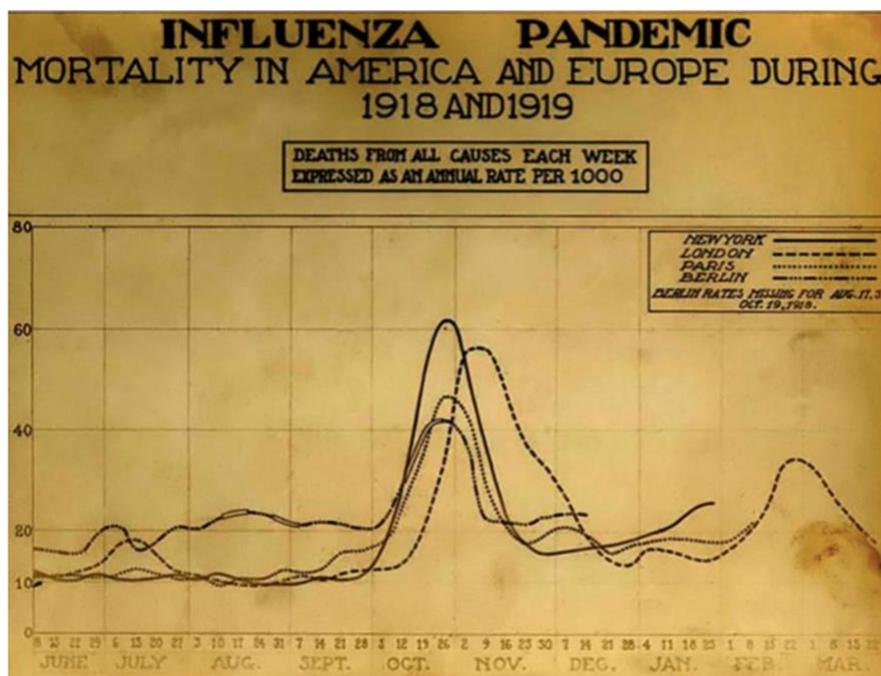
Desta forma, no presente trabalho, apesar de ser exposto diferentes tipos de distribuições, foi optado por explorar a distribuição normal devido sua grande utilização, aplicando seus conceitos no entendimento dos dados relativos à pandemia COVID-19.

### 5.1 PANDEMIAS AO LONGO DA HISTÓRIA

Ao longo da história da humanidade, ocorreram grandes epidemias e pandemias, dentre elas a gripe espanhola de 1918, cujo agente causador é o vírus H1N1, que viria a causar outra pandemia em 2009. Devido a pouca documentação nesta época, os dados não são muito precisos e estima-se que a gripe matou entre 20 e 40 milhões de pessoas.

Porém, conforme mostrado na figura abaixo, a curva da contaminação tende a se distribuir normalmente, em diferentes países.

Figura 19: Pandemia da Influenza 1918 e 1919



Fonte: NICHOLLS (2006), imagem do acervo do National Museum of Health and Medicine, Armed Forces Institute of Pathology, Washington, D.C., United States

No período de 2019 a meados de 2022, todo o mundo passou por uma nova pandemia, fruto de uma virose que ficou conhecida como COVID-19. O vírus causador pertence à família SARS, mais precisamente o SARS-CoV-2. Os sintomas no infectado são semelhantes ao da gripe comum, mas com uma maior taxa de mortalidade. Assim, com o intuito de combater essa pandemia, vários estudos foram feitos, criando diversos modelos e testes, utilizando de diferentes ferramentas da estatística, como a média móvel que foi muito utilizada para prever novos casos diários, bem como a distribuição normal.

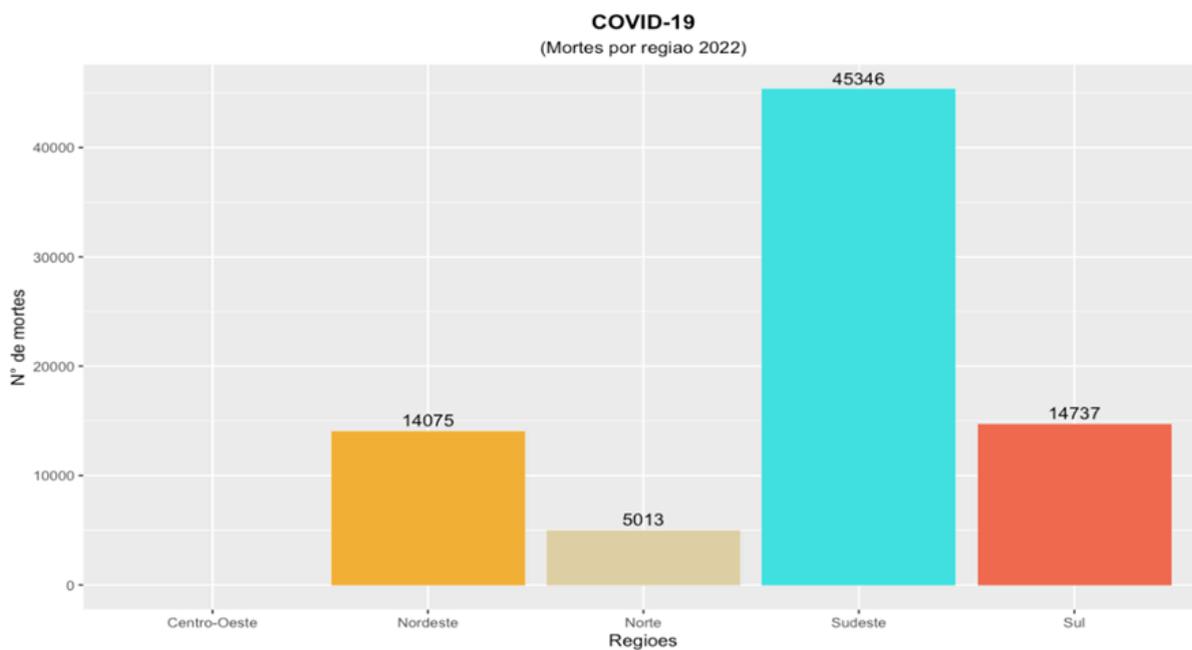
Ao longo da história é perceptível que as pandemias, em especial a gripe espanhola e a pandemia COVID19, o número de contaminados ao longo do tempo, tende a se comportar como uma normal, podendo aplicar esta técnica para prever alguns dados com relação ao futuro. Esta técnica foi amplamente usada, principalmente no início da pandemia, onde se discutia o isolamento das pessoas, como uma forma de breca a contaminação.

A seguir, será explorado algumas variáveis, bem como analisar os últimos meses do ano de 2022, utilizando a curva normal.

Os dados obtidos para estudo, foram extraídos dos sites [brasil.io](http://brasil.io) e [covid.saude.gov.br](http://covid.saude.gov.br). A fim de realizar um entendimento das variáveis, foi feita uma análise univariada e bivariada, onde as variáveis estudadas foram: data, estado, região, sexo, novos casos, novos óbitos, casos acumulados e óbitos acumulados.

Ao analisar a quantidade de mortes por região no ano de 2022, percebe-se maior concentração na região sudeste, como podemos visualizar no gráfico da figura a seguir.

Figura 20: Gráfico COVID-19 – Mortes por região



Fonte: Autoras, 2022.

No estudo dos estados do Brasil, a fim de identificar os mais afetados neste ano de 2022, com destaque para o Estado de São Paulo.

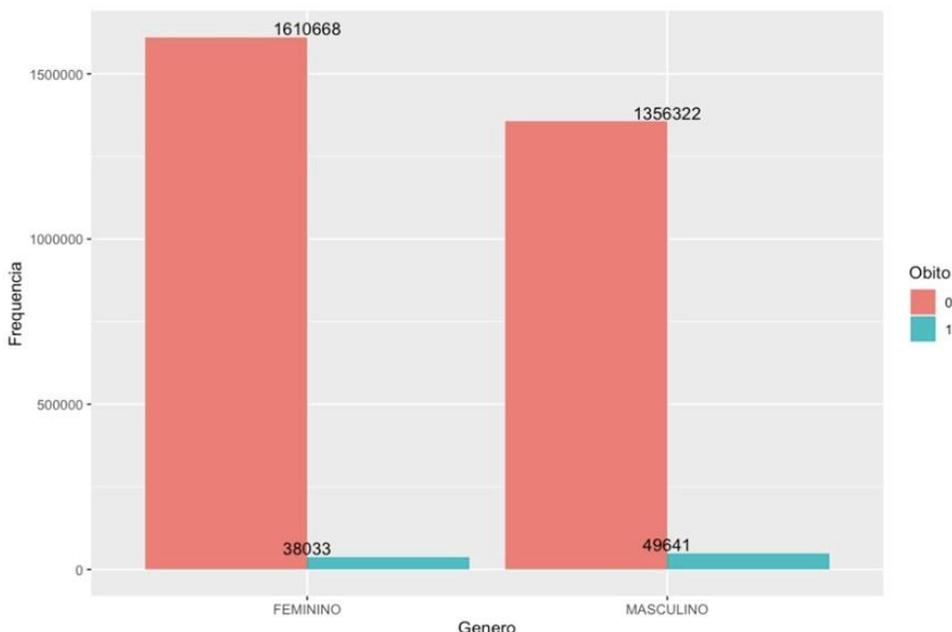
Figura 21: ESTADOS MAIS AFETADOS DO COVID-19



Fonte: Autoras, 2022.

Houve dificuldade de encontrar uma base de dados confiável, pois cada estado tem um modelo de divulgação dos dados, e no link covid.saude.gov.br o governo divulga os dados considerando apenas algumas variáveis, e como foi verificado maior concentração de casos na região sudeste em especial no Estado de São Paulo, optou-se por verificar a variável gênero, onde percebe-se maior contaminação entre as mulheres, porém maior número de óbitos entre os homens.

Figura 22: Óbitos por gênero



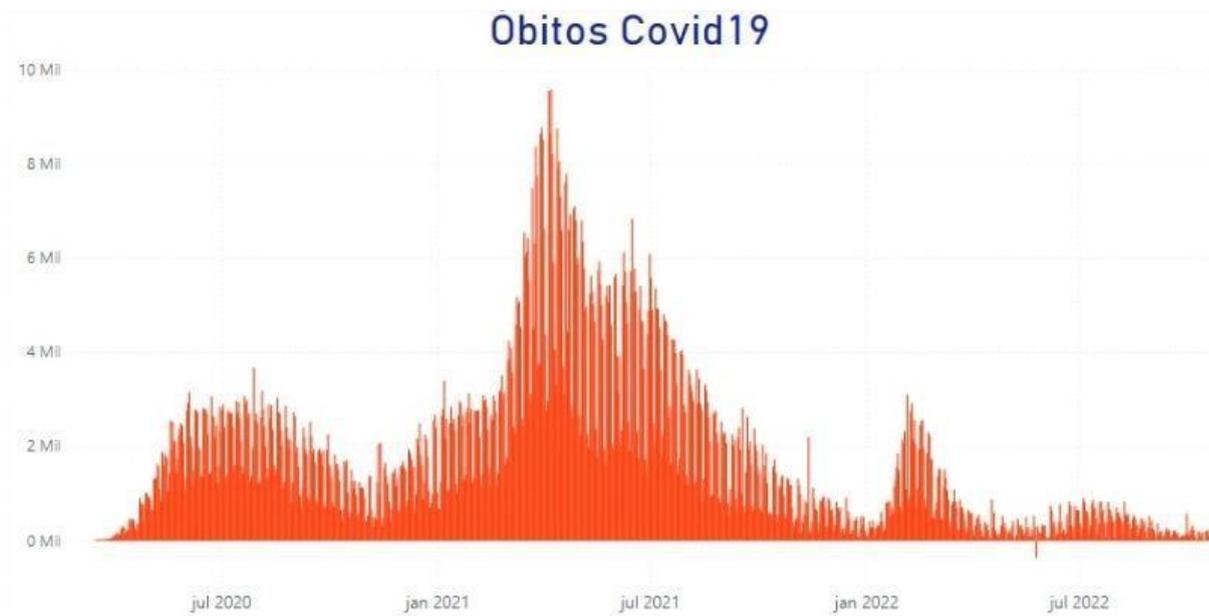
Fonte: Autoras, 2022.

Após análise de algumas variáveis, foi feito um levantamento da série histórica de contaminados pelo vírus desde 2020 no Brasil, bem como o número de óbitos, como pode-se ver nas figuras (20) e (21).

Figura 23: Evolução dos casos de COVID-19 no Brasil



Figura 24: Evolução dos casos de óbito pelo COVID-19 no Brasil

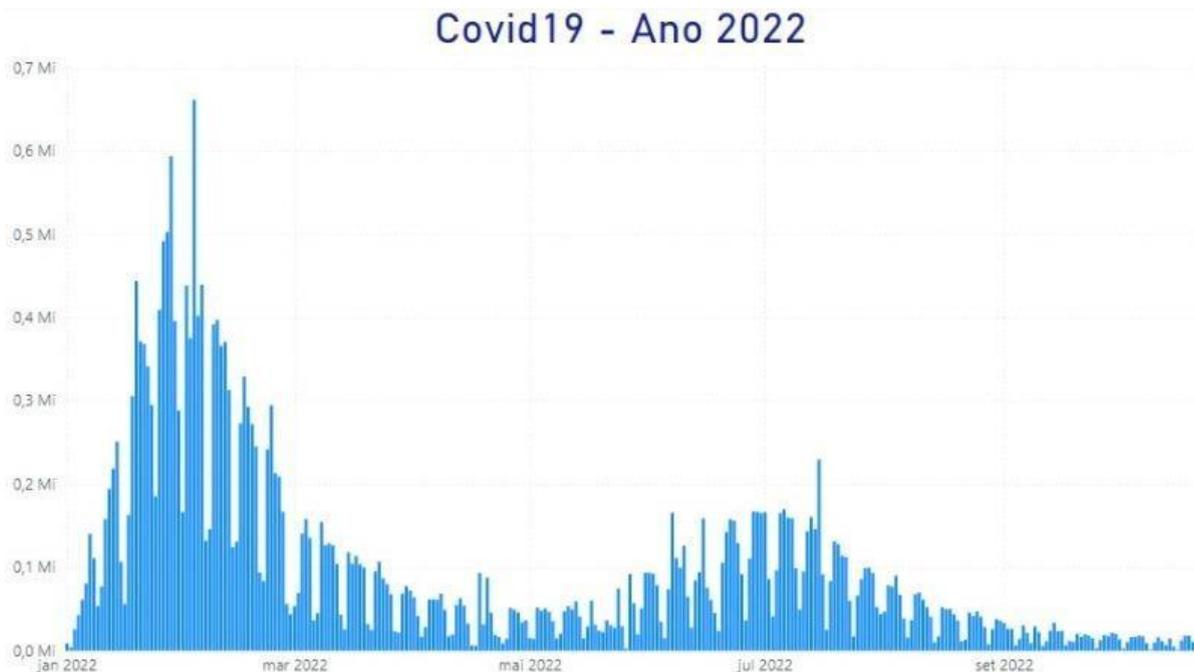


Fonte: Autoras, 2022.

Percebe-se que ao longo do tempo, a curva aproxima-se em vários momentos de uma curva normal, podendo ser aplicada para prever situações que possam ocorrer, como algum pico de contaminação, ou quando o número de casos irá diminuir, entre outros. Para isto, é preciso fazer a normalização dos dias e assim, temos os dias com um certo desvio padrão, a média que indica o dia de contaminação máxima para aquele período, dentre outros.

O gráfico abaixo, refere-se ao número de contaminados no ano de 2022, onde no primeiro trimestre do ano, ocorreu um pico de contaminação em fevereiro e em julho, apesar de ser em menor quantidade.

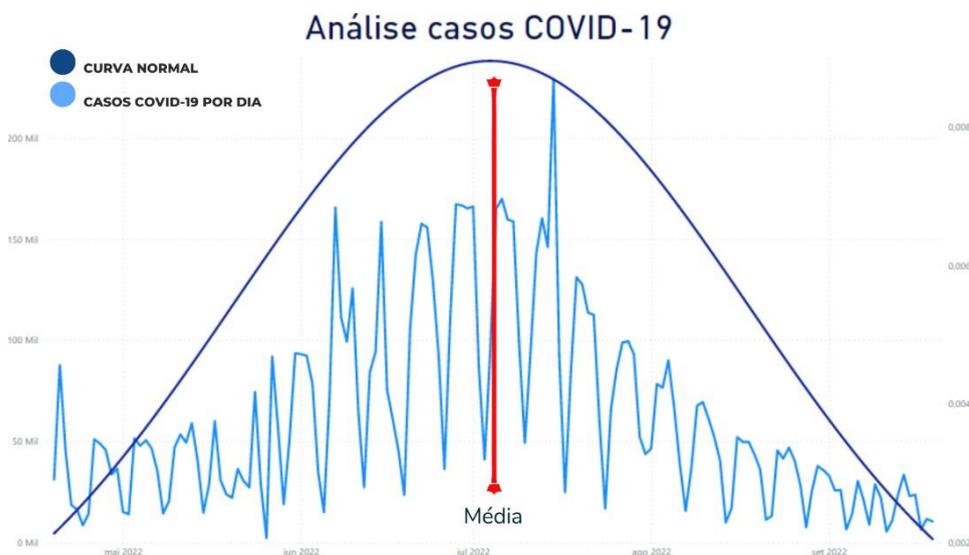
Figura 25: Contaminação entre fevereiro e julho



Fonte: Autoras, 2022.

Ao analisar a última onda de contaminação e efetuando a normalização da variável tempo, para de maio a setembro, que pode ser verificado na figura 18. Seria possível prever que o pico para este intervalo de tempo, seria por volta do dia 04/07/2022, em seguida o número de contaminação entraria em decrescimento, podendo verificar qual o percentual de contaminação para os próximos dias, tendo em vista que os dados estão concentrados em um, dois ou três desvio padrão da média.

Figura 26: Análise casos COVID-19



Fonte: Autoras, 2022.

Enfim, a curva normal pode ser explorada para diversas situações, onde temos 68% dos dados concentrados a um desvio padrão da média, 94% a dois desvios e 99,7% a três desvio, e com isto é possível tomar decisões eficazes sobre modelos que estejam sendo desenvolvidos para estudar e analisar determinado fenômeno.

O objetivo deste trabalho foi realizar uma investigação teórica sobre estatística descritiva, probabilidade e ferramentas estatísticas para processamento e análise de dados. Foram apresentados conceitos e definições, juntamente com discussões sobre medidas de centralidade e dispersão, e quatro diferentes distribuições de probabilidade, incluindo Binomial, Poisson, Exponencial e Normal.

Devido à distribuição normal ser amplamente usada, pois através dela, é possível modelar uma infinidade de fenômenos naturais, aproximar probabilidades de variáveis que possuem outras distribuições, optou-se por focar na mesma, a fim de compreendê-la e aplicá-la.

Com a aplicação da distribuição normal nas pandemias ao longo da história e dando ênfase na COVID-19, podemos perceber que a pesquisa e os gráficos da aplicação podem contribuir significativamente para a compreensão da pesquisa e análise dos dados.

## REFERÊNCIAS BIBLIOGRÁFICAS

BARBOSA, Janeide. Distribuição de Poisson. Disponível em:

<http://dspace.bc.uepb.edu.br/jspui/bitstream/123456789/8241/1/PDF%20-%20Janeide%20Alves%20Barbosa.pdf>. Acessado 10 de outubro de 2022.

BORBA, R O raciocínio combinatório na educação básica. In: Anais... 10 Encontro Nacional de Educação Matemática. Salvador-BA, 2010

BREDA, Fabiane. “Variáveis Contínuas”. Portal São Francisco, 11 de dezembro de 2020, <https://www.portalsaofrancisco.com.br/matematica/variaveis-continuas>. Acessado 08 de agosto de 2022.

COTA, Wesley. “Monitoramento do número de casos de COVID-19 no Brasil”. Wesley Cota - Physics and Complex Networks, <https://covid19br.wcota.me/>. Acessado 08 de agosto de 2022.

Covid19 por Município - Brasil.IO. <https://brasil.io/covid19/>. Acessado 17 de agosto de 2022.

DIAS, Roberto Carvalho. QUANDO ACABA?

[https://www.pucrs.br/direito/wp-content/uploads/sites/11/2020/06/2020\\_06\\_22-direito-covid-19-ppgd-documentos-quando\\_acaba-projecao\\_termino\\_fase\\_aguda\\_covid-19.pdf](https://www.pucrs.br/direito/wp-content/uploads/sites/11/2020/06/2020_06_22-direito-covid-19-ppgd-documentos-quando_acaba-projecao_termino_fase_aguda_covid-19.pdf). Acessado 12 de agosto de 2022.

Distorção | Resumo e Exemplos | Notas de revisão de matemática de nível A. <https://alevelmaths.co.uk/statistics/skewness/>. Acesso em 17 de novembro de 2022.

Distribuição Normal (Gaussiana).

<https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html>. Acessado 17 de junho de 2022.

ENCICLOPÉDIA BIOSFERA, Centro Científico Conhecer – Jandaia-GO, v.18 n.35; p. 202117 ANÁLISES ESTATÍSTICAS DO NÚMERO DE CASOS E ÓBITOS NOVOS NO ESTADO PERNAMBUCANO: SUCESSO OU FRACASSO AO COMBATE DA COVID-19? <https://www.conhecer.org.br/enciclop/2021A/analises%20estatisticas.pdf>. Acessado 08 de agosto de 2022.

FARIA, Ana Maria Lima de; KUBRUSLY, Jessica Quintanilha; SOUZA, Mariana Albi de Oliveira. GET00143 – TEORIA DAS PROBABILIDADES II : Variáveis Aleatórias Unidimensionais. Disponível em:

<<https://www.professores.uff.br/jessica/wp-content/uploads/sites/137/2017/09/GET00143-0.pdf>> . Acessado em: 04 de janeiro de 2022

FRANCO, Camila. QUAL PROJETO DEVO ESCOLHER? Ou como tomar decisões em situações de incerteza. Acessado em: 26 de setembro de 2022

HAZZAN, Samuel; Fundamentos de matemática elementar: Combinatória e Probabilidade. 5v; 8 ed; São Paulo: Atual, 2013.

HERONDINO, Prof. Distribuição normal. Disponível em:  
<https://www2.unifap.br/herondino/files/2014/04/7-DISTRIBUI%C3%87%C3%83O-NORMAL.pdf>. Acessado em: 29 de setembro de 2022

MARTINS, Maria Eugênia Graças. INTRODUÇÃO À PROBABILIDADE E À ESTATÍSTICA: Com complementos de Excel. Departamento de Estatística e Investigação Operacional da FCUL - Sociedade Portuguesa de Estatística, 2005.

MEYER, Paul L. Probabilidade Aplicações a Estatística/Paul L. MEYER; tradução Ruy de C. B. Lourenço filho. -[Reimpr.]. -Rio de Janeiro: LTC, 2010.

MOORE, Davif S. A prática da estatística empresarial como usar dados para tomar decisão. Tradução: Luís Antônio Fajardo. Rio de Janeiro: LTC, 2006.

MORGADO, A. C. O. et al. Análise combinatória e probabilidade. Rio de Janeiro: Sociedade Brasileira de Matemática, 2001. (coleção Professor de Matemática).

REIS, Elizabeth. Estatística descritiva. Lisboa: Silabo, ed. 4, 1998.  
Dados Abertos | Governo do Estado de São Paulo. *Dados Abertos | Governo do Estado de São Paulo*, <https://www.saopaulo.sp.gov.br/planosp/simi/dados-abertos/>. Acessado 10 de agosto de 2022.

RPubs - Análise COVID-19. [https://rpubs.com/Jotta\\_29/covid-19](https://rpubs.com/Jotta_29/covid-19). Acessado 17 de agosto de 2022.

Walpole, Ronald E et al. Probability & statistics for engineers & scientists/Ronald E. Walpole . . . [et al.] — 9 th. Ed. ISBN 978-0-321-62911-1. Boston-USA/2011.

## ANEXO A – TABELA DE DISTRIBUIÇÃO PADRONIZADA

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

<b>z</b>	<b>0,0</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Fonte: Autoras, 2022.